

## Ferraro Consulting FirstGlance™



As generative AI explodes on the scene, traditional data integration and API integration platforms face challenges handling massive unstructured files.

Enter DataVolo with an enterprise version of the open-source Apache NiFi, a platform created especially to help the NSA securely move and analyze the massive images, videos, and audio files necessary to keep our country safe. The code is based on NSA's "Niagara Files" software which it open-sourced as in 2014.

### Company at a Glance

DataVolo was founded in 2023 by open-source industry veterans and Cloudera alumni Joe Witt and Luke Roquet. Headquartered in Peoria, AZ, the company announced a \$21M round of funding on April 2, 2024. General Catalyst led the A round which also included Citi Ventures, Human Capital, Rob Bearden, and MVP Ventures. Given that they are building out a more complete, enterprise-ready version of opensource Apache NiFi, the company positions itself as the "leader in multimodal data pipelines for AI." In 2006, while working at the NSA, CEO, Joe Witt, was the creator of the project that became Apache NiFi. He also founded Onyara which was acquired by Hortonworks in 2015. COO, Luke Roquet, has a strong background in sales leadership at companies such as Oracle, Hortonworks, Unravel Data, AWS, and Cloudera.

### Product at a Glance

Apache NiFi automates cybersecurity, observability, event streams, and generative AI data pipelines for thousands of companies worldwide across every industry. DataVolo, powered by Apache NiFi, accelerates the creation, management, and observability of bi-directional, multimodal data pipelines for most AI use cases. DataVolo addresses the need for massive unstructured files necessary to feed AI in a timely, secure, and observable manner. The company is currently offering a beta version of their product targeting companies building Retrieval Augmented Generation (RAG) applications, seeking a SaaS solution, and looking to automate continuous capture, transformation, and loading of unstructured data.

### Ferraro Consulting POV

In the quest for [unified data engineering](#), DataVolo provides an unique advantage. It will be difficult for traditional data and API integration vendors to handle the large files that are necessary to feed both general AI and generative AI. For companies already struggling with moving, transforming, and loading data to fuel the new appetite for AI, DataVolo is a worthwhile option. In addition, as organizations are attempting to better understand lineage within the AI framework, DataVolo is already providing advanced observability including data provenance, data cleansing, data validation, data transformation, data profiling, error handling, logging, monitoring, alerts, service prioritization, and feedback loops.